Influencia de la direccionalidad de los ítem en los resultados de instrumentos de medición*

Influence of item directionality in the outcome of measurement instruments

Viviana Marycel Céspedes Cuevas**
Universidad Nacional de Colombia

Agustín Tristán-López
Instituto de Evaluación e Ingeniería Avanzada, S.C.
San Luis Potosí, México

Recibido: 3 de diciembre de 2012
Revisado: 20 de febrero de 2013
Aceptado: 30 de julio de 2013

Resumen

El trabajo ilustra los aspectos principales donde la direccionalidad de los ítems incide en la validez, la objetividad y la confiabilidad de los instrumentos de medición utilizados en educación, psicología y salud. Se trata de un elemento a tomar en cuenta en el diseño del modelo de evaluación, en la definición de las variables, la elaboración del banco de ítems para la construcción del instrumento y la forma de presentar e interpretar los reportes. La direccionalidad no depende solamente de los fundamentos teórico-conceptuales asociados con el modelo de medición y con la propia práctica (educativa o clínica, por ejemplo), sino que está regida por un comportamiento o respuesta "típica" de las personas que integran la población focal en estudio. Se muestra que la codificación de los ítems no debe hacerse en términos de "presencia-ausencia", sino del crecimiento de la variable de "poco" a "mucho" del rasgo medido en términos de respuesta "típica-atípica", "apropiada-no apropiada" u otro atributo distintivo de la variable unidimensional que puede representarse en un eje cartesiano. Se aplican los elementos del modelo de direccionalidad a un instrumento diseñado para medir la experiencia de los síntomas de mujeres con Síndrome Coronario

* Artículo de investigación.
** Correspondencia: Viviana Marycel Céspedes Cuevas, Universidad Nacional de Colombia. Dirección postal: Calle 44 No. 45-67, unidad Camilo Torres, bloque B2, oficina 705. Bogotá, Colombia. Correo electrónico: vmcespedesc@unal.edu.co. Agustín Tristán López. Dirección postal: Instituto de Evaluación e Ingeniería Avanzada, S.C. Cord. Occidental 635. Col. Lomas 4ª sección, 78216 San Luis Potosí, México. Correo electrónico: ici_kalt@yahoo.com

Agudo a partir del Modelo Conceptual del Manejo del Síntoma.

Palabras clave: cuestionario, validez, direccionalidad del constructo, experiencia de los síntomas.

Abstract

This work presents the main issues of item directionality on validity, objectivity and reliability of measurement instruments used in various fields (education, psychology or health). This is an important element to be considered in the definition of variables, the preparation of item banks, the production of tests and the interpretation of reports. Directionality depends both on the theory and the conceptual model involved in the measurement instrument, as well as the type of practice (educational or clinical for example), but it is also ruled by a "typical" behavior or response of the focal group. Item coding should not be done in terms of "presence-absence" but must take into account the direction and sense of growth of the variable as "typical-non typical", "appropriate-non appropriate" or other distinctive unidimensional trait that should be represented in a Cartesian axis. The elements of the model of directionality are used in an instrument designed to measure the experience of the women's symptoms with acute coronary syndrome based on the Symptom Management Conceptual Model.

Keywords: Questionnaire, validity, directionality of the construct, symptoms experience, woman.

Introducción

Los instrumentos semiobjetivos para la medición en educación, psicología y salud se apoyan en las respuestas que las personas ofrecen a un conjunto de ítems con los que se exploran conductas, percepciones o respuestas ante condiciones contextuales específicas. Se denominan aquí "instrumentos semiobjetivos" (IEIA, 2012) a todos los cuestionarios que recogen las opiniones o percepciones de las personas, haciéndose notar que cada ítem es planteado de forma objetiva, pero la respuesta no lo es porque depende de la apreciación de la persona que contesta. Dentro de este concepto de "instrumento semiobjetivo" se incluyen las escalas tipo Likert y cualquier otro ítem en escala categórica (por ejemplo: de nunca a siempre, de total acuerdo a total desacuerdo), las preguntas de respuesta forzada (que aparecen igualmente en el famoso artículo de Likert de 1932) y otros tipos de ítems que pueden producir resultados en puntuaciones discretas, incluyendo los de respuesta categórica (si-no, presencia-ausencia). En lo sucesivo y solo dentro de los propósitos de este documento, se denominarán "cuestionarios" o "instrumentos", sin pérdida de generalidad.

Para construir un instrumento deben identificarse las variables de interés, buscando que los valores medidos puedan representarse como puntos de un eje cartesiano cuyo crecimiento va en la dirección de la propia variable. En este trabajo se estudia la direccionalidad, como una de las propiedades de la escala que permite garantizar la validez, es dependiente de la objetividad del atributo medido e incide en la confiabilidad de las respuestas obtenidas con el instrumento. Se ilustra el impacto de la direccionalidad en un instrumento diseñado para medir la Percepción del Síndrome Coronario Agudo (SCA) en la mujer, proponiendo un modelo novedoso de direccionalidad de los ítems.

Cualidades métricas de los instrumentos

Los tres elementos principales que definen las cualidades métricas de un instrumento son: (1) atributos de medición, (2) características dimensionales y métricas de la variable, (3) propiedades de escala para la variable cartesiana (DeVellis, 2012).

La medición debe satisfacer tres atributos básicos (Tristán, 2001): (1) validez que permite medir el propósito deseado para que las conclusiones que se obtengan sirvan a dicho propósito (Borsboom, Mellenbergh & Ven Heerden, 2004); (2) objetividad que hace que el instrumento independiente de la población focal (Bond & Fox, 2007; Wright & Stone, 2004), pudiendo administrarse e interpretarse del mismo modo por una persona entrenada que por otra que no lo es, como podrían ser el profesor y sus estudiantes o el profesional y el paciente, con la salvedad de que este segundo atributo ha sido reinterpretado recientemente como manifestación de equidad por el Joint Commitee AERA, APA y NCME (2014), y (3) confiabilidad para que los resultados sean repetibles, con el mínimo error posible (Nunnally & Bernstein, 1993).

También es conveniente (aunque no obligatorio) que cada variable sea unidimensional, para que los valores puedan llegar a ordenarse en forma ascendente y ser incluyentes para que la probabilidad de respuesta sea monótona creciente en la misma dirección de la escala (Wright & Stone, 2004), que debe hacer que la variable cartesiana sea continua (cualquier valor real obtenido es interpretable dentro de la variable) y que siempre mantenga la direccionalidad con respecto del crecimiento de la variable (De Landsheere, 1976; Linacre, 2014).

Tanto el instrumento como los ítems deben diseñarse considerando las propiedades mencionadas, garantizando que los propios ítems se refieran a la misma escala de la variable, tanto si se trata de ítems de respuesta categórica, escala Likert o dicótomos (Linacre, 2014; Myers & Hansen, 2002; Wright & Stone, 2004; Wright, Stone & Enos, 2000).

Modelo de direccionalidad de ítems e instrumentos

La dirección del ítem, también llamada polaridad, se relaciona generalmente con la “presencia” o “ausencia” del rasgo medido, especialmente en el caso de los ítems dicótomos, que se codifican en dos categorías (1-0 respectivamente). Estos ítems se deben codificar tomando como referencia al eje cartesiano de la variable, de tal modo que el valor 1 en la respuesta al ítem debe corresponder con “más” del rasgo medido, en la dirección de su crecimiento o, inclusive mejor, debe corresponder con la respuesta, comportamiento o evidencia “típica” del rasgo. De este modo, el valor 0 se refiere a “menos” del rasgo medido, la respuesta corre en dirección contraria a la del crecimiento de la variable o, mejor todavía, corresponde con la respuesta, comportamiento o evidencia “atípica”. Esta codificación no es trivial, porque de su correcta aplicación depende garantizar que la probabilidad p asociada con la escala corra de 0.0 a 1.0.

Como ejemplo de esta codificación, en el caso de un cuestionario en salud, se tendría que el síntoma que ocurre más frecuentemente conviene reportarlo como respuesta “típica” con el valor 1. Igualmente si la respuesta más frecuente es la ausencia de síntoma, entonces esta respuesta sería codificada como 1, aunque pudiera ser vista como “atípica” desde el punto de vista conceptual o teórico.

Relación de la direccionalidad y los resultados en medición

Para la siguiente discusión se considera que un instrumento se administra a un grupo de personas de la población focal, cuyas medidas se ordenan de menos a más, se organiza a las personas en subgrupos de medidas (por ejemplo en este trabajo se ordenan en quintiles) y para cada ítem se determina la frecuencia de respuestas “correctas” de los quintiles. Una vez hecho esto, la direccionalidad de los ítems tiene varias consecuencias en los resultados obtenidos con el cuestionario.

Por las hipótesis del modelo conceptual de medición, se espera que el patrón de respuestas en los quintiles sea creciente.

> Con dicho patrón se puede tener un ajuste aceptable al modelo de Rasch, que proporciona una función que expresa la probabilidad de respuesta correcta de cada persona ante un ítem dado y que se representa como una curva logística monótonamente creciente (Bond & Fox, 2007; Wright & Stone, 2004).

El cambio de direccionalidad del ítem conduce a un patrón descendente o irregular y a una medida incorrecta del ítem, con el inconveniente de tener un desajuste respecto del modelo de Rasch.

El patrón esperado también se refleja en un poder de discriminación positivo que se calcula por medio de la diferencia en la proporción respuestas de los subgrupos superior e inferior. Asimismo se refleja en una correlación ítem-prueba positiva, generalmente calculada por medio de la correlación de Pearson o la correlación punto-biserial. El cambio en la direccionalidad del ítem induce que el poder de discriminación y las correlaciones sean cercanas a cero o hasta negativas (Tristán & Vidal, 2006).

El conjunto de correlaciones o las covarianzas se utilizan generalmente para obtener la matriz de varianza-covarianza empleada en el análisis factorial exploratorio (Nunnally & Bernstein, 1993). La direccionalidad incorrecta reduce las ponderaciones factoriales y se traduce en cálculos defectuosos en la identificación de los constructos obtenidos a partir de los factores principales, así como perjudica el análisis de validez discriminante con el modelo multirasgo-multimétodo (Campbell & Fiske, 1959; Wright, 1991).

Los autores del presente trabajo han hallado que la direccionalidad incorrecta también tiende a reducir las medidas de las personas y, en consecuencia, reduce las medias y las varianzas de los puntajes en la variable en estudio (Tristán & Céspedes, 2013). Además de que se reduce el valor numérico de la medida de los ítems, sin embargo, los autores destacan que la varianza del ítem permanece constante independientemente de la direccionalidad. A partir de este comportamiento en las varianzas de los puntajes de las personas y de los ítems, la direccionalidad incorrecta reduce la consistencia interna que puede estimarse a través del coeficiente alfa de Cronbach (Cronbach, 1951).

Aplicación del modelo en un instrumento de medición en salud

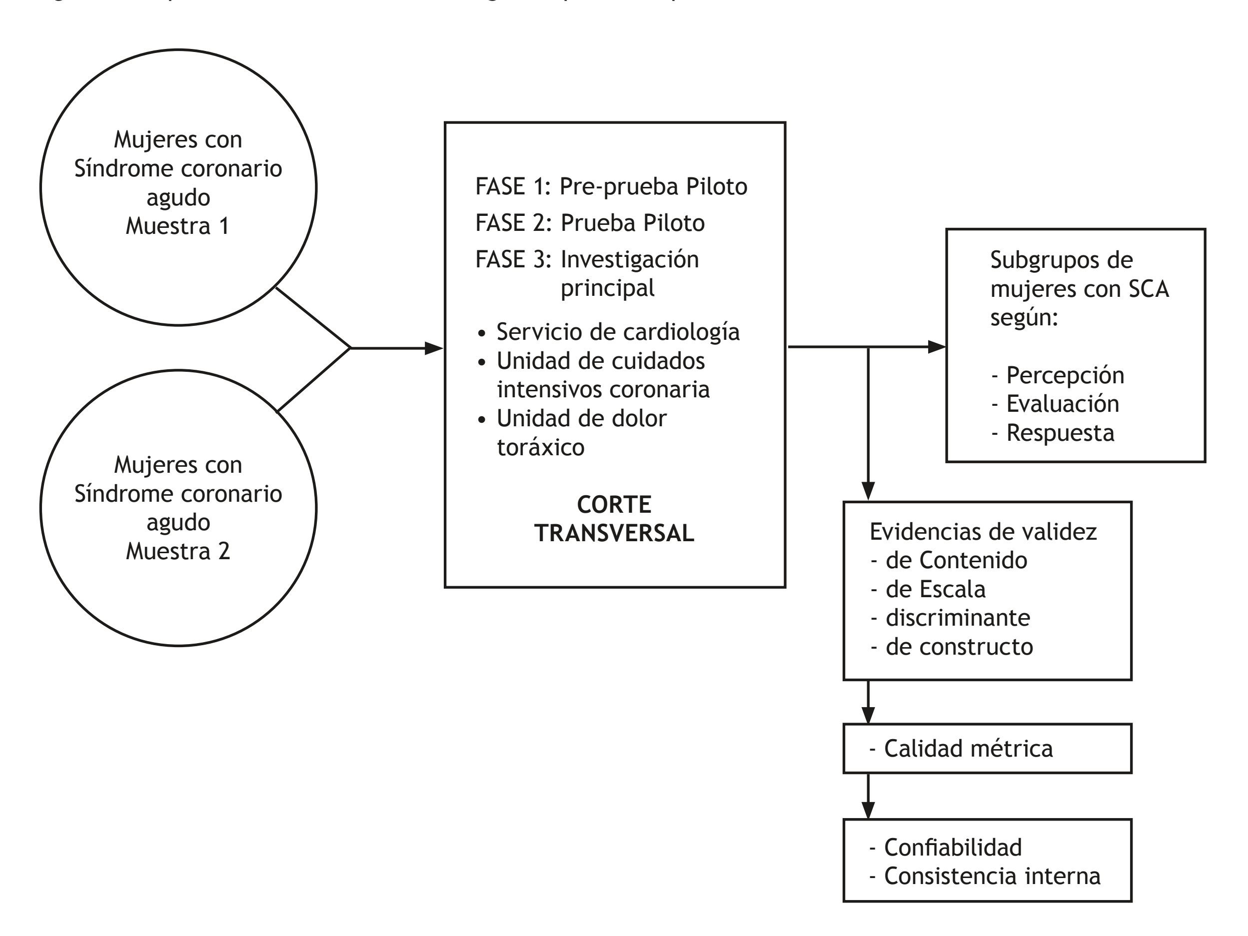
Los puntos relevantes de la sección anterior, los cuales plantean las bases del modelo de direccionalidad de los ítems propuesto por los autores, se aplican en un instrumento de medición de Síndrome Coronario Agudo (SCA) en mujeres. Para caracterizar el síntoma en el SCA se cuenta con varios modelos sobre los factores que influyen en la patología, tanto clínicos, cognitivos como de patrones sociales del síntoma. Se partió del Modelo Conceptual del Manejo del Síntoma-SMM (Caldwell & Miaskowski, 2000; Dodd, 2001; University of California San Francisco School of Nursing, 1994) que considera que las personas experimentan síntomas en dimensiones interrelacionadas e influidas por variables contextuales internas (personales) y externas (ambientales y del proceso de salud-enfermedad). Las tres dimensiones son: (1) la Experiencia del síntoma con tres constructos (percepción individual del síntoma, evaluación de su significado y respuesta al mismo), (2) las Estrategias de manejo que incluye acciones que influyen o controlan la experiencia del síntoma, previniéndolo o retrasando su impacto y (3) los Resultados que se obtienen por la intervención de elementos como habilidades de autocuidado, estatus socioeconómico, emocional y funcional, entre otros (Céspedes, 2009).

> El estudio se interesa en el diseño de un instrumento para medir las tres variables de la Experiencia del síntoma atendiendo a las características de medición indicadas previamente, con lo que se cubre un vacío en la literatura donde se presentan numerosos instrumentos de diseño incompleto o de interpretación dudosa que no garantizan el cum-

plimiento de los atributos de la evaluación (Céspedes, 2009).

Para el diseño del instrumento se desarrolló un trabajo cuantitativo descriptivo y exploratorio de corte transversal, realizado sobre una muestra de 380 mujeres mayores de 20 años, con diagnóstico confirmado de Síndrome Coronario Agudo (lesiones mayores del 75 % indicativas de enfermedad epicárdica por angiografía coronaria), con estado de conciencia y orientación conservados, con comunicación verbal conservada y hemodinámicamente estables, distribuidas en dos centros hospitalarios de Bogotá, Colombia (Figura 1).

Figura 1. Esquema del diseño de la investigación para la Experiencia del síntoma.



Se utilizó un muestreo intencional con determinación de tamaño muestral en función de la fracción de desviación estándar según el Modelo modificado de determinación de tamaño de muestra en población finita (Tristán, 2008), cuyo error de aproximación para estudio definitivo correspondió a 0.05.

El instrumento se construyó sobre la base de un conjunto de hipótesis de índole causal sobre la experiencia del síntoma dividida en los tres constructos citados (Amaya & Céspedes, 2008; Céspedes, 2009):

Constructo 1: Percepción (33 ítems). Se divide en síntomas típicos sobre dolor retro-esternal opresivo irradiado a brazos, cuello o mandíbula, o no irradiado y síntomas atípicos acerca de dolor en otras partes del cuerpo, palpitaciones, síntomas digestivos, respiratorios, músculo-esqueléticos, vegetativos, emocionales, cognitivos y aquellos relacionados con el control de la temperatura corporal.

Constructo 2: Evaluación (14 ítems). Permite identificar el origen cardíaco, no asociado con la patología, o psicosomático, atendiendo a su apreciación de seriedad como insignificante, dañino o amenazante.

Constructo 3: Respuesta (29 ítems). Distingue entre acciones o actividades que realiza la paciente u otras personas cercanas ante los síntomas. También considera el tiempo de demora para atender su caso, en función del lapso transcurrido desde que se perciben los síntomas hasta recibir tratamiento.

Direccionalidad de los ítems en el instrumento

La objetividad del instrumento se apoya en ítems de respuesta dicotómica, inicialmente codificados como 0 (ausencia) y 1 (presencia). El análisis de los resultados condujo a concluir que dicha codificación no se asocia con la dirección de crecimiento de la variable medida, por lo que se vio la conveniencia de ajustar la interpretación del binomio "presencia-ausencia" a los conceptos "típico-atípico", "asociado-no asociado a la patología cardíaca", "serio-no serio", dependiendo en cada caso de las necesidades de la dimensión a medir, de acuerdo con lo indicado en el modelo de direccionalidad, para garantizar que la medida correspondiera con el crecimiento de la variable, de esta manera:

Percepción. Se considera la percepción "típica" si la persona percibe un síntoma de tipo cardíaco, básicamente opresión en el pecho, en tanto que la percepción "atípica" es cuando la persona percibe síntomas en otras partes de su cuerpo. Bajo este enfoque el eje de esta dimensión también podría denominarse PERCEPCIÓN CARDIACA → NO CARDIACA. La codificación de estos ítems se hace "típica = 0" y "atípica = 1", la razón de que se haya definido el eje en esta forma es que la percepción atípica es predominantemente femenina, por lo que también pudo haberse denominado: PERCEPCIÓN MASCULINA → FEMENINA. Esta fue la variable que requirió mayor atención porque en la codificación preliminar (presencia=1, ausencia=0), podía revelar la percepción como cardíaca y no cardíaca pero no reflejar la respuesta "típica" o "atípica" desde el punto de vista clínico que tiene un enfoque predominantemente masculino, es decir, sesgo por género.

Evaluación. La evaluación inadecuada es cuando la persona considera que no es un problema cardíaco y, por lo tanto, no es de importancia, lo cual se codifica como 0. La evaluación adecuada es cuando la persona se da cuenta que es un problema cardíaco, que representa un riesgo que puede ser grave e importante, codificada como 1.

Respuesta. Se tiene una respuesta inadecuada codificada como 0, cuando la persona hace uso de remedios caseros, espera a que pase la molestia y, en consecuencia, el tiempo de demora para atenderse es largo. Una respuesta adecuada es cuando se atiende de inmediato, con apoyo en un profesional médico particular o en un hospital, codificada como 1.

A continuación se presentan cuatro ítems para ejemplificar la codificación y su forma de interpretación.

Ítem 1) ¿Sintió dolor o molestia en la mitad del pecho? SI-NO [Una respuesta SI (presencia) se codifica como 0, para referirse a una percepción típica muy frecuente en los hombres, pero atípica en las mujeres que son la población focal y cuyas respuestas definen el sentido de crecimiento de la escala].

Ítem 13) ¿Tuvo usted una sensación de indigestión? SI-NO [La respuesta SI (presencia) se codifica como 1, por tratarse de la percepción de un síntoma atípico de mayor frecuencia en las mujeres].

Ítem 34) ¿Pensó que la causa de lo que estaba sintiendo era un problema cardíaco? SI-NO [En este caso, la respuesta SI (presencia) se codifica como 1 y permite reportar una evaluación "adecuada", en la dirección esperada hacia el Síndrome Coronario en el estudio].

Ítem 38) ¿Pensó que su dolor era de origen dental? SI-NO [Una respuesta SI (presencia) se debe codificar 0, porque indica una evaluación "inadecuada" en dirección no esperada hacia el Síndrome Coronario en el estudio].

Los ítems se calibran con las respuestas codificadas de las personas de la población focal, verificando que cada variable debe satisfacer la definición de su direccionalidad. La verificación de la direccionalidad se hace en función de la frecuencia total de respuestas y del patrón de las respuestas en los quintiles.

Análisis de resultados

Los análisis presentados a continuación incluyen seis elementos, los tres primeros son a nivel del ítem, uno es la combinación ítem-variable y los dos últimos se relacionan con el cuestionario en su conjunto: (1) la revisión de los patrones de respuesta, (2) la discriminación con el modelo clásico; (3) el ajuste al modelo de Rasch, (4) la correlación ítem-variable, (5) el análisis factorial exploratorio y (6) la consistencia interna. Para ello se utilizaron los programas Winsteps®, Kalt® y SPSS®.

Los dos primeros análisis se reportan simultáneamente en la Tabla 1 que presenta los patrones de los quintiles y la relación discriminativa de seis de los ítems de la prueba obtenidos con el programa Kalt® (IEIA, 2014). El patrón correcto para la direccionalidad implica que la frecuencia de respuesta correcta sea pequeña en el quintil 1 y se incremente en forma monótona hasta la mayor frecuencia en el quintil 5. Obsérvese que cuando no se toma en cuenta la direccionalidad de los ítems en general, no se tiene el comportamiento creciente esperado. La Tabla 1 también presenta la relación discriminativa RD, parámetro que se espera que sea igual o superior a 1 si el ítem distingue apropiadamente entre las personas de medida más alta y las de medida baja. Son inconvenientes los valores de RD inferiores a 0.3 o negativos.

Tabla 1.
Patrones de respuesta y discriminación en los ítems del instrumento.

Ítem	Frecuencias por quintil y relación discriminativa					
	Sin direccionalidad	%	RD	Con direccionalidad	%	RD
	1	5.3		1	14.2	
	2	2.9		2	16.8	
1	3	2.1	-1.63	3	18.2	0.82
	4	1.3		4	19.2	
	5	0.8		5	19.2	
	1	8.9		1	0.0	
	2	6.8		2	7.1	
34	3	13.4	0.19	3	7.4	2.14
	4	11.8		4	18.9	
	5	7.6		5	18.2	
	1	19.7		1	0.0	
	2	13.4		2	0.8	
44	3	5.0	-1.2	3	15.0	2.33
	4	5.0		4	18.2	
	5	9.5		5	18.7	
	1	12.9		1	2.9	
	2	12.6		2	2.4	
55	3	15.0	0.45	3	5.0	1.45
	4	16.0		4	6.1	
	5	19.5		5	7.9	
	1	4.2		1	11.1	
	2	2.9		2	14.7	
57	3	3.7	0.53	3	14.5	0.76
	4	5.3		4	18.9	
	5	5.3		5	19.5	
	1	11.3		1	2.6	
	2	12.9		2	3.4	
58	3	14.5	0.33	3	4.7	1.67
	4	13.7		4	4.7	
	5	17.1		5	15.0	

En el ítem 1 de la Tabla 1 se presentan muy pocas respuestas que además son más frecuentes en el quintil más bajo produciendo una RD negativa. Obsérvese que al asignar la dirección correcta se tienen frecuencias altas en todos los quintiles con una tendencia creciente muy clara y RD cercana al valor deseable de 1. El ítem 34 sin codificación no tiene un patrón claro, con baja RD=0.19, en cambio con la direccionalidad apropiada se tiene una configuración creciente más evidente y una alta RD=2.14. Cuando el ítem 44 no cuenta con la direccionalidad bien definida se presenta un patrón de discriminación negativa, en cambio queda muy claro su patrón correcto con la codificación apropiada y una muy alta RD=2.33. Los ítems 55 y 58 hacen evidente que no se trata solamente de tener una codificación que brinde altas frecuencias de respuesta, sino que reflejen la direccionalidad. La asignación apropiada de la direccionalidad hace que la relación discriminativa pase de 0.45 a 1.45 en el ítem 55 y de 0.33 a 1.67 en el ítem 58.

Al estudiar el ítem 2 que explora el dolor o molestia en la mitad del pecho, opresivo irradiado al cuello, brazos o mandíbula, se observa (Tabla 2) que no presenta una tendencia creciente en ninguna de las dos codificaciones direccionales (RD=0.35 sin direccionalidad y RD=0.04 con direccionalidad), lo cual puede indicar que se trata de un ítem no apropiado para las mujeres, aunque es reconocido como un elemento "típico" en los hombres.

Tabla 2.
Patrones de respuesta y discriminación en el ítem 2.

Ítem	Sin direccionalidad: quintil	%	RD	Con direccionalidad: quintil	%	RD
2	1	2.9	0.35	1	15.5	0.04
	2	5.5		2	14.5	
	3	6.1		3	14.2	
	4	4.7		4	15.0	
	5	8.4		5	13.2	

La siguiente fase del análisis combina el ajuste al modelo logístico y la correlación punto-biserial que se realiza por medio de la calibración con el modelo de Rasch (Tabla 3A) usando el programa Winsteps® (Linacre, 2014). Este software produce la medida de cada ítem (MEASURE) y los parámetros de ajuste (INFIT y OUTFIT) reportados con los modelos de media cuadrática (MNSQ) y Z estandarizada (ZSTD). Los valores aceptables de MNSQ deben ser menores a +1.3 y de ZSTD deben ser menores a +2.0. El mismo programa proporciona la correlación punto-biserial (rpbis) en la columna PTMEA CORR cuyo valor de aceptación debe ser superior a 0.196 (Tristán y Vidal, 2006).

Un error en la codificación de la direccionalidad incide en defectos respecto del modelo logístico. Sin tomar en cuenta la direccionalidad se tienen 17 ítems que desajustan por INFIT, OUTFIT (señalados con asterisco) o por valores bajos o negativos de rpbis, pero estos resultados aparentemente inconvenientes cambian al codificar correctamente la direccionalidad. Se observó que los ítems 1 y 44, entre otros, ya no presentan una correlación negativa y el ítem 2 cambia su correlación punto biserial a un valor negativo, y el ítem 55 tiene un valor rpbis=0.09, en cambio, solo los ítems 3 y 48 desajustan de manera importante en INFIT ZSTD con valores superiores a 2.0.

Tabla 3.

Resultados del análisis de Rasch SIN direccionalidad. Ítems ordenados de mayor a menor desajuste (OUTFIT ZSTD)

ENTRY NUMBER Número de ítem	TOTAL SCORE Suma de puntos	TOTAL COUNT Total de respuestas	MEASURE Medida	MODEL S.E. Error estándar	INFIT		OUTFIT		PTMEA CORR rpbis
					MNSQ	ZSTD	MNSQ	ZSTD	
44	200	380	-0.30	0.11	1.25*	7.9*	1.28*	7.3*	-0.09
1	47	380	1.88	0.16	1.18	1.5	1.98*	5.3*	-0.23
34	185	380	-0.13	0.11	1.14	4.8*	1.19	5.0*	0.06
46	180	380	-0.07	0.11	1.11	3.8*	1.17	4.7*	0.10
3	201	380	-0.31	0.11	1.12	3.8*	1.17	4.6*	0.10
45	181	380	-0.08	0.11	1.11	3.7*	1.17	4.6*	0.10
57	264	380	-1.06	0.11	1.20	3.9*	1.30	4.3*	-0.05
54	289	380	-1.42	0.12	1.18	2.6*	1.29	3.2*	-0.03
36	43	380	1.98	0.16	1.02	0.2	1.39*	2.3*	0.09
76	154	380	0.23	0.11	1.07	2.2*	1.10	2.3*	0.16
56	81	380	1.20	0.13	1.11	1.6	1.20	2.0	0.04
49	146	380	0.32	0.11	1.07	2.1*	1.08	1.8	0.17
50	55	380	1.69	0.15	1.05	0.6	1.23	1.7	0.07
47	279	380	-1.27	0.12	1.09	1.6	1.11	1.5	0.12
55	30	380	2.39	0.19	1.06	0.4	1.27	1.3	0.03
2	105	380	0.84	0.12	1.04	0.8	1.09	1.2	0.18
31	9	380	3.66	0.33	0.95	0.0	1.11	0.4	0.13

(*) = desajuste al modelo de Rasch

Fuente: Adaptada de la salida de Winsteps®

Tabla 3B

Resultados del análisis de Rasch CON direccionalidad. Ítems ordenados de mayor a menor desajuste (OUTFIT ZSTD)

ENTRY NUMBER Número de ítem	TOTAL SCORE Suma de puntos	TOTAL COUNT Total de respuestas	MEASURE Medida	MODEL S.E. Error estándar	INFIT		OUTFIT		PTMEA CORR rpbis
					MNSQ	ZSTD	MNSQ	ZSTD	
2	275	380	-1.04	0.12	1.25	4.2*	1.68*	7.1*	-0.08
3	201	380	-0.12	0.11	1.10	3.1*	1.21	4.6*	0.17
23	129	380	0.74	0.11	1.05	1.1	1.13	2.1*	0.23
48	146	380	0.53	0.11	1.10	2.7*	1.11	2.1*	0.18
36	43	380	2.23	0.16	0.99	0.0	1.34*	1.9	0.19
56	299	380	-1.41	0.13	1.09	1.2	1.18	1.7	0.14
54	92	380	1.25	0.12	1.10	1.5	1.16	1.7	0.15
55	350	380	-2.62	0.19	1.05	0.4	1.25	1.1	0.09
40	33	380	2.54	0.18	1.00	0.0	1.17	0.9	0.19
50	55	380	1.93	0.15	1.03	0.4	1.12	0.9	0.18

(*) = desajuste al modelo de Rasch

Fuente: Adaptada de la salida de Winsteps®

Los demás ítems reportados en la Tabla 3B presentan valores de desajuste de menor importancia en la medición, con valores cercanos a los límites aceptables de los parámetros de ajuste.

El análisis de validez discriminante puede realizarse con base en la correlación punto-biserial, con estos pasos: (1) Un comité de expertos revisa la definición teórica de cada variable; (2) se identifican los ítems que corresponden con la variable de acuerdo con el criterio de jueces o expertos; (3) se ordenan los ítems en función de la correlación ítem-prueba calculada con la fórmula de Pearson o como correlación punto-biserial; (4) se acepta el conjunto de ítems que tienen altas correlaciones positivas ítem-prueba y (5) se eliminan los ítems que correlacionan poco o negativamente. Cabe resaltar que cada uno de los conjuntos de ítems, se considera como una posible variable si se satisface que: 1. La variable cuenta con cuatro ítems por lo menos (se puede considerar una variable con tres ítems si se obtienen valores altos en las siguientes dos condiciones), 2. Los ítems correlacionan positivamente con el conjunto (valores superiores a 0.196) y 3. El conjunto produce valores altos del coeficiente de consistencia interna alfa. Siguiendo este procedimiento es posible reorganizar las tres dimensiones del cuestionario en constructos internamente consistentes (Tabla 4) con interpretación clínica clara.

Tabla 4.
Agrupaciones por dimensiones y correlación ítem-total.

Dimensión	Subdimensión	Número de ítems en la subdimensión	Ítems
Percepcion	Atípica mixta	31	1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 32, 33
Evaluación	Origen extracardiaco mixto	10	34, 35, 36, 37, 38, 39, 40, 41, 42, 43
	Seriedad insignificante	4	44, 45, 46, 76.
Respuesta	Inadecuada	8	47, 49, 51, 52, 53, 55, 56, 57

La quinta fase consiste en el análisis factorial exploratorio del instrumento, que se realizó con los ítems codificados sin direccionalidad y con direccionalidad, con el método de rotación varimax usando el software SPSS. Con la codificación tomando correctamente la direccionalidad de las variables se pueden distinguir ocho factores interpretables, organizados en tres de percepción atípica (mixta, digestiva y emocional), tres de origen extra cardíaco (dos mixtos y uno emocional) y dos de respuesta inadecuada.

Desde el punto de vista teórico se observó que la codificación inapropiada de los ítems hace que la matriz de varianza-covarianza se construya sobre datos incorrectos que, a su vez, producen resultados factoriales carentes de interpretación teórica conceptual o clínica (Tabla 5A y 5B). Por ejemplo la evaluación de origen extracardiaco se separa en dos factores mixtos con 7 ítems en total, en contraste con la codificación usando la direccionalidad correcta que permite identificar más ítems que intervienen en factores mixtos, y además aparece otro factor que corresponde con una evaluación de origen emocional por parte del paciente.

Tabla 5A.
Agrupaciones para percepción por análisis factorial SIN direccionalidad.

Dimensión	Subdimensión	Número de ítems en la subdimensión	Ítems
Percepción atípica	Mixta	15	4, 6, 8, 9, 11, 14, 15, 17, 18, 20, 21, 27, 28, 29, 30
	Digestiva	7	13, 22, 23, 24, 25, 26, 32
Evaluación origen extracardiaco	Mixto (1)	5	34, 44, 45, 46, 76
	Mixto (2)	2	35, 40
Respuesta	Inadecuada (1)	3	51, 52, 53
	Inadecuada (2)	2	47, 49

Tabla 5B.
Agrupaciones por análisis factorial CON direccionalidad.

Dimensión	Subdimensión	Número de ítems en la subdimensión	Ítems
Percepción Atípica	Mixta	13	1, 4, 5, 6, 8, 9, 11, 14, 15, 17, 20, 21, 30
	Digestiva	8	12, 13, 22, 23, 24, 25, 26, 32
	Emocional	5	3, 18, 27, 28, 29
Evaluación origen extracardiaco	Mixto (1)	4	44, 45, 46, 76
	Emocional	3	41, 42, 43
	Mixto (2)	7	34, 35, 36, 37, 38, 39, 40
Respuesta	Inadecuada (1)	3	51, 52, 53
	Inadecuada (2)	6	47, 49, 50, 55, 56, 57

Este análisis de validez de constructo realizado con un modelo factorial exploratorio se complementa con un análisis para demostrar que las escalas involucradas en el instrumento son unidimensionales, pudiendo aplicarse en cuestionarios por separado o, como en este caso, integrados en un solo instrumento.

La última fase del análisis se refiere a la consistencia interna del instrumento, reportada generalmente con el coeficiente alfa de Cronbach. Desde el punto de vista teórico, al revisarse la fórmula del coeficiente alfa se puede concluir que la direccionalidad correcta incide en la mejora del valor del coeficiente de consistencia interna alfa de Cronbach (Tabla 6). El cambio que experimenta alfa al cambiarse la codificación de la direccionalidad de los ítems puede ser pequeño en variables con más de 30 ítems, como en el caso de la dimensión de percepción que sube de 0.801 a 0.807, pero puede ser muy notable especialmente en variables de pocos ítems, como en la dimensión “Evaluación” que pasa de 0.048 a 0.799.

Tabla 6.
Coeficientes de consistencia interna global y por dimensión.

Dimensión	Número de ítems	Sin direccionalidad			Con direccionalidad		
		Medida % promedio	Desv. Estándar %	Alfa	Medida % promedio	Desv. Estándar %	Alfa
Global	76	43.24	8.62	0.697	45.92	9.71	0.764
Percepción	33	54.16	16.59	0.801	57.79	16.82	0.807
Evaluación	14	28.10	11.52	0.048	30.41	22.20	0.799
Respuesta	11	57.05	16.79	0.386	57.82	19.03	0.607

Discusión de resultados

Se mostró que la direccionalidad de los ítems, requiere que la codificación de las respuestas en cada ítem se haga de acuerdo con el crecimiento de la variable en un eje cartesiano que corre de “poco” a “mucho”, asociando los valores (1-0) a las manifestaciones o respuestas “típica-atípica”, “aceptable-no aceptable”, “seria-no seria” del constructo en estudio. Al establecer la direccionalidad de cada ítem se tiene un impacto en la puntuación global de respuestas y en la medida logística del ítem, pero principalmente el énfasis recae en varios elementos métricos: el patrón monótono creciente de respuestas en subgrupos (organizados por ejemplo en quintiles), un mejor ajuste al modelo de Rasch, incremento en el poder de discriminación y en la correlación ítem-prueba de cada una de las dimensiones, factores o constructos que intervienen en un instrumento dado.

Un ítem mal codificado produce valores negativos en la correlación ítem-prueba y en la discriminación, desajustando al modelo de Rasch. Al emitir conclusiones con un ítem mal direccionado se alteran la objetividad, la validez de escala y de constructo, por lo que puede llegarse a identificar factores de interpretación dudosa. En resumen, se inducen errores en las propiedades métricas del instrumento y resultados incorrectos al aplicarlos a una persona, induciendo interpretaciones inapropiadas, como por ejemplo un mayor error de calificación en los estudiantes o deficiencias en las intervenciones clínicas en los pacientes.

Conclusiones

Se mostraron los elementos descriptivos del modelo teórico de direccionalidad de los ítems para cualquier instrumento que se aplique en educación, salud o psicología. Se aplica el modelo al diseño de un instrumento de medición de la Experiencia del síntoma del SCA en mujeres y los tres constructos involucrados: Percepción, Evaluación y Respuesta.

El instrumento desarrollado y revisado con la metodología propuesta mejora la medición de la Experiencia del síntoma de SCA en mujeres. En particular se mejoró la dimensión de Percepción y se pudo diseñar el instrumento que mide las dimensiones de Evaluación y de Respuesta, que no habían sido estudiadas satisfactoriamente en la literatura. El modelo permite obtener evidencias que apoyan la validez, la objetividad y la confiabilidad del instrumento, reforzando los modelos teóricos, analíticos y estadísticos disponibles en la literatura (Céspedes, 2009). El análisis de validez de la direccionalidad, junto con el análisis de la escala, permite disponer de tres escalas unidimensionales en un mismo instrumento, las cuales pueden ser aplicadas tanto de forma independiente como en su versión completa.

El énfasis del presente trabajo es mostrar que la direccionalidad de los ítems soporta la validez de la escala de la variable y que un error de dirección induce a notables defectos en la medición. Lo más relevante es evitar las conclusiones erróneas acerca

de las variables y dictámenes incorrectos o deficientes a partir de los resultados de la medición.

El modelo de direccionalidad de los ítems, propuesto y desarrollado por los autores, expone la importancia de la temática y sus implicaciones en la medición en general y en salud en particular, con énfasis en un fenómeno específico para enfermería cardiovascular. Las implicaciones de la direccionalidad y las posibles aplicaciones para el diseño y análisis de instrumentos son diversas e importantes. Como extensión de este estudio, se contempla la validez de escala de un instrumento, con implicaciones en varias facetas, desde la construcción de un cuestionario, la adaptación y traducción de instrumentos publicados y utilizados en otros países, la verificación de la dimensionalidad de las variables y las propiedades métricas de cada escala, entre otros elementos.

Referencias

Amaya, P. & Céspedes, V. (2008). Síntoma(s) en la mujer con enfermedad coronaria: Exploración de concepto para la práctica clínica y la investigación desde la disciplina de enfermería. Enfermería Cardiovascular (pp. 147-160).

American Educational Research Association, American Psychological Association y National Council on Measurement in Education (2014). Standards for Educational and Psychological Testing. Washington: AERA.

Bond, T.G. & Fox, C.M. (2007). *Applying the Rasch model. 2a ed.* Erlbaum, NJ. (pp. 4-8).

Borsboom, D., Mellenbergh G. J. & Ven Heerden, J. (2004). *The concept of validity.* Psychological Review, *111*(4), 1.061-1.071.

Caldwell, M. & Miaskowski, C. (2000). *The Symptom experience of angina in women.* Pain management nursing, *1*(3), 69-78.

Campbell, D. T. & Fiske, D. W. (1959). *Convergent and Discriminant Validation by the Multitrait -Multimethod Matrix.* Psychological Bulletin, *56*, 81-105.

Céspedes, V. (2009). *Modelo Conceptual del Manejo del síntoma: clasificación por percepción, evaluación y respuesta de mujeres con síndrome coronario agudo; originada por la construcción de un instrumento validado en Bogotá, Colombia.* Universidad Nacional de Colombia. Tesis doctoral.

Cronbach, L. J. (1951). *Coefficient alpha and the internal structure of the tests.* Psychometrika, *16*, 297-334.

De Landsheere, G. (1976). *Introduction a la recherche en éducation.* Paris: Armand Colin-Bourrelier.

DeVellis, R. F. (2012). *Scale development.* Theory and Applications (Applied Social Research Methods). California: SAGE Publications Inc. Thousand Oaks.

Dodd, M., Janson, S., Facione, N., Faucett, J., Froelicher, E., Humphreys, J., & et al. (2001). *Advancing the science of symptom management.* J. Advance Nurs. *33*, 668-676.

Instituto de Evaluación e Ingeniería Avanzada (2014). *Manual de usuario Kalt Plus.* México: IEIA, www.ieia.com.mx

Instituto de Evaluación e Ingeniería Avanzada (2012). *Manual de usuario Kalt Opinion, para cuestionarios semiobjetivos.* México: IEIA, www.ieia.com.mx

Linacre, J. M. (2014). *A users's guide to WINSTEPS.* Recuperado de www.winsteps.org.

Myers, A. y Hansen, C. H. (2002). *Experimental psychology.* 5th edition. California: Wadsworth, Thompson Learning.

Nunnally, J. C. & Bernstein, I. H. (1993). *Psychometric Theory.* 3a *ed.* New York: McGraw-Hill.

Tristán, L. A. (2001). *Análisis de Rasch para todos.* México: Ceneval.

Tristán, L. A. & Céspedes, C. V. M. (2013). *Análisis de las variaciones en parámetros psicométri-*

cos por efecto de la direccionalidad de los ítems. Reporte interno. Proyecto N. 130512. México: Instituto de Evaluación e Ingeniería Avanzada, S.C.

Tristán, L. A. & Vidal, U. R. (2006). *Manual de fórmulas de correlación*. Canadá: Trafford. Victoria.

Tristán, L. A. (2003). *Cálculo del tamaño de muestra y establecimiento de criterios y el problema del tamaño de la población*. Guía de usuario Kalt Versión 2. Instituto de Evaluación e Ingeniería Avanzada, San Luis Potosí, México.

University of California (1994). *A model for symptom management*. San Francisco School of Nursing Symptom Management Faculty Group. J. Nurs Scholarship, *26*, 272-276.

Wright, B. (1991). *Factor Analysis versus Rasch Analysis of Items*, Rasch Measurement Transactions, *5*(1), 134-135.

Wright, B. D. & Stone M. H. (2004). *Making measures*. The Phaneron Press. Chicago. USA. 35-39.

Wright, B. D., Stone, M.H. & Enos, M. (2000). *Evolution of meaning in practice*. Rasch Measurement Transactions, *14*(1), 736.